

Clone Your Voice Without Sounding Like a Robot

The clone isn't the hard part. The script is. Here's how to set one up, write for it, and catch it when it quietly skips your words.

Here's what nobody tells you about voice clones.

Making one takes about 2 minutes. Making one that doesn't sound like a robot reading a letter... that's where everybody gets stuck, and they blame the wrong thing.

They think the clone is broken. They re-record their samples. They pay for the expensive version. And it still sounds flat.

The clone was never the problem. The script was.

This is everything I use, in the order you need it.

Step 1: Pick the right kind of clone

ElevenLabs gives you 2 doors, and picking the wrong one wastes a week.

Instant clone. You upload about 1 to 2 minutes of clean audio of yourself talking. It's ready almost straight away. ElevenLabs' own docs say don't go past 3 minutes, more audio doesn't help and can make it worse.

Professional clone. You upload at least 30 minutes, and they recommend 2 to 3 hours. Then it trains, which can take anywhere from a few hours to a full day. It needs the Creator plan or higher.

Start with instant. Almost everybody should. Use it for a month. Only go professional when you can name the specific thing instant won't do for you.

I'll be straight with you, I trained the professional one. It started slipping into random accents halfway through sentences. I went back to instant and never looked back. More audio isn't automatically a better clone.

Step 2: Record samples that don't sabotage you

The clone copies *everything*. Your pitch, your pace, your breathing, and also the echo in your kitchen and the fridge humming behind you.

- **Quiet room, no echo.** A closet full of clothes beats a nice empty office.

- **1 speaker.** Nobody else talking, no TV.
- **Talk the way you want it to sound.** If you read your samples slow and careful, you get a slow careful clone. Talk like you're explaining something to a friend.
- **Same energy the whole way through.** Don't start hyped and finish tired.

Step 3: Make your first line

Text to Speech. Your voice selected on the right. Type a sentence. Generate.

That's the whole setup. Play it back and let it sit for a second, because the first time is weird.

The honest part: every generation spends credits. Check your usage page after the first few so it's not a surprise later.

Step 4: Fix the robot (this is the part that matters)

Your first line will be fine. It'll also be *flat*. Flat is what makes people say "that sounds like AI."

You fix it in the script, not the settings. 3 tools:

Audio tags

Put a direction in square brackets and the voice performs it:

```
[sighs] Look... I tried to do this the hard way for months. [laughs] Don't be me.
```

Tags that work: [laughs], [sighs], [whispers], [excited], [curious], [sarcastic]. You're not changing *what* it says. You're telling it *how* to say it.

In my test, the same sentence went from 6 seconds flat to 9.7 seconds with tags in it. That extra 3.7 seconds IS the performance.

Tags only work on the newer models. If tags get read out loud or ignored, check which model you've got selected before you blame the script.

Punctuation is direction

- A full stop is a breath.
- An ellipsis is a pause... and it adds weight.
- CAPITALS add emphasis. Use them on 1 word, not a whole sentence.

Write it the way you'd say it

This is what everybody writes first:

I wanted to reach out because I noticed that your current setup is costing you a significant amount of time each week.

Reads fine. Sounds like a robot reading a letter. Now cut it up:

Your setup is costing you hours. Every week... [sighs] and there's a better way.

Same meaning. Short pieces. A pause where you'd trail off. A tag where you'd sigh. That's the version that stops sounding generated.

The rule: you're not writing prose. You're writing a script for a performer, and the punctuation is your stage directions.

Step 5: The stability slider

Stability runs from Creative on 1 end to Robust on the other.

- **Toward Creative:** more expressive, follows your tags harder, occasionally goes a bit wild.
- **Toward Robust:** steady and consistent, but flatter.

If your tags aren't landing, slide it toward Creative and generate again. If it's going off the rails, slide it back.

Step 6: The check nobody does

This one cost me. A long script came back **missing 32 words**. No error. A perfectly valid file that played start to finish... with 10 seconds of dead air where a whole chunk of the script should've been.

It only got caught because somebody listened to the whole thing.

On long text, assume it skipped something until you've proved it didn't.

- **Chunk it.** Keep each generation short enough that you'd notice a gap. A few paragraphs at a time, not the whole script.
- **Read along while it plays.** Not skim. Read. It's the only way to catch it.
- **Every time,** on anything long. A file that plays is not a file that's correct.

The 1 minute checklist

Before you post anything made with a clone:

1. Instant clone, 1 to 2 minutes of clean audio.

2. Script written in short pieces, the way you talk.
3. Tags and ellipses where you'd pause or react.
4. Newer model selected, stability nudged toward Creative if it's flat.
5. Long scripts generated in chunks.
6. Listened to every second while reading along.

That's the whole playbook. The people selling voice cloning courses are selling you step 1. Steps 4 through 6 are where it starts sounding like you.

There's a lot more where this came from

Everything I give away is free and lives at jordybaze.com/free. Start with whichever of these fits what you're stuck on:

- **Make the video without the crew:** jordybaze.com/free/video
- **The scam that uses your kid's voice:** jordybaze.com/free/scams
- **Content without it eating your whole week:** jordybaze.com/free/content

Free from jordybaze.com/free/voice · updated 2026-10-27.